****
*** TX REPORT ***
****

TRANSMISSION OK

TX/RX NO	3310
CONNECTION TEL	917035185499
SUBADDRESS	
CONNECTION ID	
ST. TIME	10/22 13:14
USAGE T	03'06
PGS.	9
RESULT	OK

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Fax Cover Sheet

Date: 22 Oct 2008

To: Benjamin J. Hauptman (29, 310)	**From:** MICHAEL PHAM
Application/Control Number: 10/791,897	**Art Unit:** 2167
Fax No.: 703-518-5499	**Phone No.:** (571)272-3924
Voice No.: ~~970-000-0000~~ 703-884-1111	**Return Fax No.:** (571) 273-8300
Re:	**CC:**

☐ **Urgent** ☒ **For Review** ☐ **For Comment** ☐ **For Reply** ☐ **Per Your Request**

Comments:
Attatched are proposed amendments to improve clairity and to put the case into condition for allowance. Please let me know by Friday, October 24, 2008, 10am eastern time if authorization for an examiner's amendment can be done, such that the case may be completed.

Number of pages 9 including this page

Draft Amendments:

Claim 1 :

A method of clustering documents each having one or plural document segments in an input document set, said method comprising the following steps:

(a) obtaining a co-occurrence matrix for each input document which is a matrix reflecting the occurrence frequencies of terms and the co-occurrence frequencies of term pairs, and obtaining an input document frequency matrix for the set of input documents based on occurrence frequencies of terms or term pairs appearing in the set of input documents wherein said step (a) further includes:

(a-l) generating an input document segment vector for each of said input document segments based on occurrence frequencies of terms appearing in each input document segment;

(a-2) obtaining the co-occurrence matrix for each input document from the input document segment vectors; and

(a-3) obtaining the input document frequency matrix from the co-occurrence matrix for each document;

(b) selecting a seed document from a set of remaining documents that are not included in any cluster existing at that moment, and constructing a current cluster of an initial state based on the seed document, wherein said selecting and constructing comprises:

(b-1) constructing a remaining document common co-occurrence matrix for the set of the remaining documents based on a product of corresponding

components of the co-occurrence matrices of all documents in the set of remaining documents; and

(b-2) obtaining a document commonality of each remaining document to the set of the remaining documents based on a product sum between every component of the co-occurrence matrix of each remaining document and the corresponding component of the remaining document common co-occurrence matrix;

(b-3) extracting, as the seed document, the document having the highest document commonality to the set of the remaining documents; and

(b-4) constructing the initial cluster by including the seed document and neighbor documents similar to the seed document;

(c) making documents, which have the document commonality to the current cluster higher than a threshold, belong temporarily to the current cluster; wherein said making comprising:

(c-l) constructing a current cluster common co-occurrence matrix for the current cluster and a current cluster document frequency matrix of the current cluster based on occurrence frequencies of terms or term pairs appearing in the documents of the current cluster;

(c-2) obtaining a distinctiveness value of each term and each term pair for the current cluster by comparing the input document frequency matrix with the current cluster document frequency matrix;

(c-3) obtaining weights of each term and each term pair from their distinctiveness values;

(c-4) obtaining a document commonality to the current cluster for each document in the input document set based on a product sum between every component of the co-occurrence matrix of the input document and the corresponding component of the current cluster common co-occurrence matrix while applying the respective weights to said components; and

(c-5) making documents having the document commonality to the current cluster higher than the threshold belong temporarily to the current cluster;

(d) repeating step (c) until the number of documents temporarily belonging to the current cluster does not increase;

(e) repeating steps (b) through (d) until a given convergence condition is satisfied;

and

(f) deciding, on the basis of the document commonality of each document to each cluster, a cluster to which each document belongs and outputting said cluster.

Claim 2 (canceled)

Claim 5 :

The clustering method according to claim 1, wherein the convergence condition in said step (e) is satisfied when

(i) the number of documents whose document commonalities to any current clusters are less than a threshold becomes 0, or

(ii) the number is less than a threshold and does not increase.

Claim 6 :

The clustering method according to claim 1, wherein said step (f) further includes: checking existence of a redundant cluster, and removing, when the redundant cluster exists, the redundant cluster and again deciding the cluster to which each document belongs.

Claim 7 :

A method of clustering documents each having one or plural document segments in an input document set, said method comprising the following steps:

(a) obtaining a co-occurrence matrix S^r for each input document D_r based on occurrence frequencies of terms or term pairs appearing in the set of input documents;

wherein in step (a), each mn component S^r_{mn} of the co-occurrence matrix S^r of the document D_r is determined in accordance with:

$$S^r_{mn} = \sum_{y=1}^{Y_r} d_{rym} d_{ryn}$$

where:

m and n denote m^{th} and n^{th} terms, respectively, among M terms appearing in the set of input documents,

D_r is the r^{th} document in a document set D consisting of R documents;

Y_r is the number of document segments in document D_r, wherein d_{rym} and d_{ryn} denote the existence or absence of the m^{th} and n^{th} terms, respectively, in the y^{th} document segment of document D_r, and

S^r_{mm} represents the number of document segments in which the m^{th} term occurs and S^r_{mn}

represents the co-occurrence counts of document segments in which the m^{th} and n^{th} terms co-occur;

(b) selecting a seed document from a set of remaining documents that are not included in any cluster existing at that moment, and constructing a current cluster of an initial state based on the seed document, wherein said selecting and constructing comprise:

(b-1) constructing a remaining document common co-occurrence matrix T^A for the set of the remaining documents based on the co-occurrence matrices of all documents in the set of remaining documents;

(b-2) obtaining a document commonality of each remaining document to the set of the remaining documents based on the co-occurrence matrix S^r of each remaining document and the remaining document common co-occurrence matrix T^A;

(b-3) extracting, as the seed document, the document having the highest document commonality to the set of the remaining documents; and

(b-4) constructing the initial cluster by including the seed document and neighbor documents similar to the seed document;

(c) making documents having the document commonality higher than a threshold belong temporarily to the current cluster;

(d) repeating step (c) until the number of documents temporarily belonging to the current cluster does not increase;

(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and

(f) deciding, on the basis of the document commonality of each document to each cluster, a cluster to which each document belongs and outputting said cluster.

Claim 9 :

The method according to claim 7, wherein in step (b-1), the remaining document common co-occurrence matrix T^A is determined on the basis of a matrix T;

wherein the matrix T has an mn component determined by

$T_{mn} = \prod_{r=1}^{R} S^r_{mn}$ and

$S^r_{mn} > 0$

the matrix T^A has an mn component determined by

$T^A_{mn} = T_{mn}$ when $U_{mn} > A$,

$T^A_{mn} = 0$ otherwise,

where

U_{mn} represents an mn component of a document frequency matrix of the set of remaining documents wherein U_{mm} denotes the number of remaining documents in which the m^{th} term occurs and U_{mn} denotes the number of remaining documents in which the m^{th} and n^{th} terms co-occur; and

A denotes a predetermined threshold.

Claim 10:

The method according to claim 9, further comprising:

determining a modified common co-occurrence matrix Q^A on the basis of T^A ; and in step (b-2), obtaining the document commonality of each remaining document to the set

of the remaining documents based on the co-occurrence matrix Sr of each remaining document and the modified common co-occurrence matrix Q^A;

the matrix Q^A having an mn component determined by

$Q^A{}_{mn} = \log T^A{}_{mn}$ when $T^A{}_{mn} > 1$,

$Q^A{}_{mn} = 0$ otherwise.

Claim 11 :

The method according to claim 10, wherein in step (b-2),

the document commonality of each remaining document P having a co-occurrence matrix S^P with respect to the set of remaining documents is given by

$$com_q(D', P; Q^A) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{M} Q^A_{mn} S^P_{mn}}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M}(Q^A_{mn})^2}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M}(S^P_{mn})^2}}.$$

Claim 12 :

The method according to claim 10, wherein in step (b-2), the document commonality of each remaining document P having a co-occurrence matrix S^P with respect to the set of remaining documents is given by

$$com_q(D', P; Q^A) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{M} T^A_{mn} S^P_{mn}}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M}(T^A_{mn})^2}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{M}(S^P_{mn})^2}}.$$

Claims 23-24 (canceled)

Claims 27-28 (canceled)

Claims 29-31 (canceled)